Letter to the Editor

Better limited systematicity in hand than structural descriptions in the bush: A reply to Hummel

Hummel's reference to "countless other ways" in which our model of structure processing is deficient, along with the claim that it cannot account for "any of the hallmarks of the human ability to perceive spatial relations," exemplifies the tactic favored by the defenders of the structural description (SD) approach across cognition: broaden the scope of the discussion to reduce the impact of a refutation of a particular instance of the theory. A similar situation prevails in linguistics, where no amount of evidence seems to suffice to send transformational generative grammar to the dumpster: the theory of grammar keeps mutating over the decades (Edelman & Christiansen, 2003), and there is always some "performance-related" excuse for sticking to rule-based explanations. This will not do: the proponents of SD should decide whether or not their theory is refutable in principle in the context of a well-defined set of observations, such as those offered in (Hummel, 2000) and a well-defined model, such as that of (Hummel, 2001) or (Hummel & Biederman, 1992).

Refuting the SD approach was not, however, our goal in the article commented upon by Hummel. Likewise, the ability of SD models to deal with any entities standing in any relation to each other is not ignored by us (as Hummel claims in the penultimate paragraph). On the contrary, it is precisely this property of SD that makes them too powerful (hence biologically irrelevant, unless properly constrained), as we discuss at length in Section 1.2. Consequently, we set out to develop a limited, yet conceptually and computationally plausible alternative to SD, to implement it using biologically uncontroversial mechanisms—a category that includes what + where neurons, but excludes dynamic binding (Lennie, 1998; Kirschfeld, 1995)—and to test the implementation on real, manually unpreprocessed images (something that no SD model is capable of at present). In this admittedly circumscribed project we have succeeded.

We note that Hummel's equation of what + where with conjunctive coding betrays a basic misunderstanding of our approach. It is the combination of what + where code at one level with a bottom-up/top-down "computation cube" scheme that enables our model to exhibit (limited) systematicity. This computational stance leads to predictions of analogous limitations in human performance, which we state in Section 5.4. We also note that the crucial psychophysical experiments have yet to be performed; we are aware of no published studies of the representation of structure that use unfamiliar-looking or even just complex shapes and configurations while discouraging scrutiny and abstract, extra-visual reasoning on the part of the subject.

In summary, we believe (appropriating Mahatma Gandhi's quip concerning the Western civilization) that SD would be a good idea—(1) if the visual world were uniquely describable in terms of compositions of parts, (2) if human observers were systematic in their processing of novel arrangements of unfamiliar visual stimuli, and (3) if the detection of the parts and relations were computationally feasible. Unfortunately, the stance expressed by the first condition is philosophically naive (Akins, 1996; Smith, 2001), the second is empirically unfounded, and the third is increasingly perceived by the computer vision community as unrealistic (Barrow & Tenenbaum, 1993; Dickinson et al., 1997). It looks like limited systematicity may prove to be the only game in town.

References

Akins, K. (1996). Of sensory systems and the 'aboutness' of mental states. *Journal of Philosophy*, *XCIII*, 337–372.

Barrow, H. G., & Tenenbaum, J. M. (1993). Retrospective on "Interpreting line drawings as three-dimensional surfaces". *Artificial Intelligence*, *59*, 71–80.

Dickinson, S., Bergevin, R., Biederman, I., Eklundh, J., Munck-Fairwood, R., Jain, A., & Pentland, A. (1997). Panel report: The potential of geons for generic 3-d object recognition. *Image and Vision Computing*, *15*, 277–292.

Edelman, S., & Christiansen, M. H. (2003). How seriously should we take minimalist syntax? A comment on Lasnik. *Trends in Cognitive Science*, *7*(2), 60–61.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition*, *8*, 489–517.

Hummel, J. E. (2000). Where view-based theories of human object recognition break down: The role of structure in human shape perception. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (Chap. 7). Hillsdale, NJ: Erlbaum.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.

Kirschfeld, K., 1995. Neuronal oscillations and synchronized activity in the central nervous system: Functional aspects. *Psycoloquy 6*(36). Available electronically as: ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1995.volume.6/psyc.95.6.36.brain-rhythms.11.kirschfeld

Lennie, P. (1998). Single units and visual cortical organization. *Perception*, *27*, 889–936.

Smith, B. (2001). Fiat objects. *Topoi*, *20*, 131–148.

Shimon Edelman*
Department of Psychology, Cornell University
232 Uris Hall, Ithaca, NY 14853-7601, USA
Nathan Intrator
School of Computer Science, Tel Aviv University
Ramat Aviv, Tel Aviv 69978, Israel
* Corresponding author. Tel.: +1-607-255-6365
fax: +1-607-255-8433
E-mail address: se37@cornell.edu